

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-265396

(43)公開日 平成11年(1999) 9月28日

(51)Int.Cl.<sup>9</sup>

識別記号

F I

G 0 6 F 17/30

G 0 6 F 15/40

3 7 0 D

G 1 0 L 9/00

G 1 0 L 9/00

A

G 0 6 F 15/401

3 1 0 D

審査請求 未請求 請求項の数9 O L (全 5 頁)

(21)出願番号 特願平10-68159

(22)出願日 平成10年(1998) 3月18日

(71)出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72)発明者 南 遼一

東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(72)発明者 阿久津 明人

東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(72)発明者 佐藤 隆

東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(74)代理人 弁理士 小笠原 吉義 (外1名)

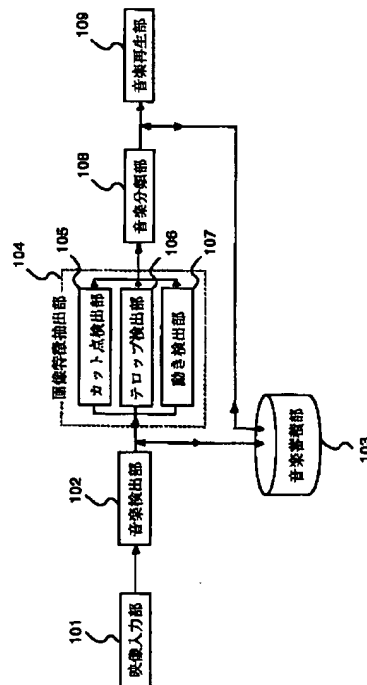
最終頁に続く

(54)【発明の名称】 音楽映像分類方法、装置および音楽映像分類プログラムを記録した記録媒体

(57)【要約】

【課題】 本発明は、音楽映像の画像に含まれる特徴を用いて、曲のテンポの早さを自動的に検出し、音楽を分類することを目的としている。

【解決手段】 映像情報を入力し、映像情報の音情報から音楽を検出し、検出された音楽に対応する画像情報の特徴を抽出し、画像の特徴に基づいて音楽を分類する。



## 【特許請求の範囲】

【請求項1】 与えられた映像情報から音楽映像を分類する方法であって、

映像情報を入力する映像入力段階と、

映像情報における音情報から音楽を検出する音楽検出段階と、

検出された音楽を蓄積する音楽蓄積段階と、

検出された音楽に対応する画像情報の特徴を抽出する画像特徴抽出段階と、

前記画像の特徴に基づいて音楽を分類する音楽分類段階と、

蓄積された音楽を検索、再生する音楽再生段階と、

を実行することを特徴とする音楽映像分類方法。

【請求項2】 前記画像特徴抽出段階において、場面の替わり目であるカット点の有無、テロップの有無および大きさ、被写体あるいはカメラの動きのうち、少なくとも1つを検出することを特徴とする請求項1記載の音楽映像分類方法。

【請求項3】 前記音楽分類段階は、カット点の単位時間あたりの出現頻度、テロップの単位時間あたりの出現頻度を大きさに正規化した値、被写体あるいはカメラの動きの大きさのうち少なくとも1つに基づいて音楽を分類することを特徴とする請求項2記載の音楽映像分類方法。

【請求項4】 与えられた映像情報から音楽映像を分類する装置であって、

映像情報を入力する映像入力部と、

映像情報における音情報から音楽を検出する音楽検出部と、

検出された音楽を蓄積する音楽蓄積部と、

検出された音楽に対応する画像情報の特徴を抽出する画像特徴抽出部と、

前記画像の特徴に基づいて音楽を分類する音楽分類部と、

蓄積された音楽を検索、再生する音楽再生部と、

を具備することを特徴とする音楽映像分類装置。

【請求項5】 前記画像特徴抽出部は、場面の替わり目であるカット点の有無を検出するカット点検出部、テロップの有無および大きさを検出するテロップ検出部、被写体あるいはカメラの動きを検出する動き検出部のうち、少なくとも1つを具備することを特徴とする請求項4記載の音楽映像分類装置。

【請求項6】 前記音楽分類部は、カット点の単位時間あたりの出現頻度、テロップの単位時間あたりの出現頻度を大きさに正規化した値、被写体あるいはカメラの動きの大きさのうち少なくとも1つに基づいて音楽を分類することを特徴とする請求項5記載の音楽映像分類装置。

【請求項7】 与えられた映像情報から音楽映像を分類するプログラムを記録した記録媒体であって、

映像情報を入力する映像入力処理と、

映像情報における音情報から音楽を検出する音楽検出処理と、

検出された音楽を蓄積する音楽蓄積処理と、

検出された音楽に対応する画像情報の特徴を抽出する画像特徴抽出処理と、

前記画像の特徴に基づいて音楽を分類する音楽分類処理と、

蓄積された音楽を検索、再生する音楽再生処理と、

をコンピュータに実行させるための音楽映像分類プログラムを記録したことを特徴とする記録媒体。

【請求項8】 前記画像特徴抽出処理は、場面の替わり目であるカット点の有無を検出するカット点検出処理、テロップの有無および大きさを検出するテロップ検出処理、被写体あるいはカメラの動きを検出する動き検出処理のうち、少なくとも1つを実行するようにしていることを特徴とする請求項7記載の記録媒体。

【請求項9】 前記音楽分類処理は、カット点の単位時間あたりの出現頻度、テロップの単位時間あたりの出現頻度を大きさに正規化した値、被写体あるいはカメラの動きの大きさのうち少なくとも1つに基づいて音楽を分類するようにしていることを特徴とする請求項8記載の記録媒体。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、画像の特徴を用いて音楽映像を分類する音楽映像分類方法、装置および音楽映像分類プログラムを記録した記録媒体に関する。

【0002】

【従来の技術】音楽はポップス、ジャズ、ロック、演歌などのジャンルの他、悲しい曲、楽しい曲、テンポの早い曲、遅い曲などの分類があるが、自動的にこのような分類を行うためには高度な音楽認識技術が必要であり、実用には至っていない。ただテンポの抽出には、ドラムのビートを周波数スペクトルの急激な立ち上がりから検出し、利用するものがある。

【0003】

【発明が解決しようとする課題】ドラムのビートを検出する方法では、ドラムのような強いパワーを持つ楽器を用いない場合には、正確なテンポが抽出出来ないという問題があった。

【0004】本発明は、音楽映像の画像に含まれる特徴を用いて、曲のテンポの早さを自動的に検出し、音楽を分類することを目的としている。

【0005】

【課題を解決するための手段】上記目的を達成するため、本発明においては、映像情報を入力し、映像情報の音情報から音楽を検出し、検出された音楽に対応する画像情報の特徴を抽出し、画像の特徴に基づいて音楽を分類することにより、ビートが明確に検出できないような

音楽でもテンポによる分類ができるようにしている。

【0006】また、場面の替わり目であるカット点の有無、テロップの有無および大きさ、被写体あるいはカメラの動きのうち、少なくとも1つを検出することにより、音情報を用いずに音楽のテンポを算出することができるようにしている。

【0007】また、カット点の単位時間あたりの出現頻度、テロップの単位時間あたりの出現頻度を大きさと正規化した値、被写体あるいはカメラの動きの大きさのうち少なくとも1つに基づいて音楽を分類するようにしている。

【0008】

【発明の実施の形態】以下に、本発明の実施例について図面を参照して説明する。図1は、本発明の一実施形態の音楽映像分類装置の概略構成を示すブロック図である。本実施形態の音楽映像分類装置は、映像を入力する映像入力部101と、音楽を検出する音楽検出部102と、検出された音楽を蓄積する音楽蓄積部103と、画像の特徴を抽出する画像特徴抽出部104と、音楽を分類する音楽分類部108と、蓄積された音楽を検索し、再生する音楽再生部109から構成されている。また、画像特徴抽出部104は、カット点検出部105と、テロップ検出部106と、動き検出部107から構成されている。

【0009】図2は、本発明の一実施例の音楽映像分類装置の処理の流れを示したフローチャートである。本発明をソフトウェアで実現した場合でも同様の処理の流れとなる。

【0010】まず、音楽検出処理201において、映像における音情報が解析され、音楽の検出が行われる。音楽の検出は、音情報の周波数スペクトルのピークが、周波数方向に対して時間的に安定しているという特徴を用いる方法（南他、「音解析による映像インデクシング」、電子情報通信学会総大会、D-12-64、1997）などを用いることが可能である。

【0011】なお当該「音解析による映像インデクシング」は、映像に含まれる音情報から、音声や音楽を自動的に検出し、これらが含まれる部分のみを抜き出して映像を要約するものである。例えば、歌番組のトークを聞かずに歌の部分のみを聞きたいといった場合に有効である。音楽が存在する場合、周波数スペクトルのピークは、周波数方向に対して時間的に安定しているという特徴があることから、ピークを検出し、時間的な持続性を算出することによって、音楽を検出することができる。

【0012】判断202において音楽の開始が判断され、音楽が開始された場合には、音楽蓄積処理203によって音楽を蓄積しながら、対応する画像情報から画像特徴抽出処理204が行われる。画像特徴抽出処理204では、カット点の検出、テロップ領域の検出、被写体あるいはカメラの動きの検出のうち少なくともいずれか

1つが行われる。カット点の検出には、画素値のフレーム間差分を用いる方法（谷口他、「映像ショット切換え検出法とその映像アクセスインタフェースへの応用」、電子情報通信学会論文誌J79-D-2、No. 4、1996）などが利用可能である。

【0013】なお当該「映像ショット切換え検出法とその映像アクセスインタフェースへの応用」は、映像の場面の変わり目であるカット点を自動的に検出し、カットの直後の静止面を順次並べて表示することによって、映像の内容が一覧できる。この文献の方法では、映像のフレーム間差分を求めることによって、カット点を求めることを基本とし、隣接するフレームだけでなく、少し間をおいたフレームとの差分を評価することによって、フェード等のゆっくりとした変化も捉えることができる。また、応用例として、1日分のテレビの映像の一覧を作成し、好きな時間の映像を見ることができるシステムを構築している。

【0014】また、テロップの検出には、テロップの領域が背景部分に比べて高輝度であり、エッジを抽出しやすいという性質を用いた方法（茂木他、「ニュース映像中の文字認識に基づく記事の索引付け」、電子情報通信学会技術研究報告IE95-153、1996）や、映像符号化の国際標準であるMPEGなどのフレーム間の相関を用いて符号化された映像データのうち、フレーム間の相関を用い、かつ動き補償を用いないで符号化された画素が時間的・空間的に集中している部分を抽出する方法（特願平8-131898「動画像テロップ抽出方法及び装置」）などの方法を用いることができる。

【0015】なお前記「ニュース映像中の文字認識に基づく記事の索引付け」は、映像中に現れる文字情報（テロップ）は、映像の検索に有効であり、ニュースを例に取ると、ニュースの項目や主要な場面でテロップが現れる。テロップの出現を自動的に検出することによって、映像をニュース項目ごとに分割したり、テロップが表示されている画像を並べるだけで大体的内容を知ることが可能である。この文献では、テロップの領域が背景部分に比べて高輝度であり、エッジを抽出しやすいという性質を用いた方法を取り入れている。

【0016】また前記「動画像テロップ抽出方法及び装置」は、映像符号化の国際標準であるMPEGなどのフレーム間の相関を用いて符号化された映像データのうち、フレーム間の相関を用い、かつ動き補償を用いないで符号化された画素が時間的・空間的に集中している部分を抽出する方法である。目的は、ニュース映像中の文字認識に基づく記事の索引付けと同じである。

【0017】さらに、被写体あるいはカメラの動きを検出する方法については、動きベクトルの特徴を用いる方法（阿久津他、「動画像インデクシング目的としたカメラ操作の規定方法」、電子情報通信学会論文誌J75-D-2、No. 2、1992）などが利用可能である。

【0018】なお当該「動画像インデクシング目的としたカメラ操作の規定方法」は、映像を撮影する時には、カメラを横に振ったり、ズームしたりといったことが行われるが、これは、空間的な広がり表現するためや、ある特定の箇所をより詳しく見せるために行うものである。よって、そのような操作を伴う映像は、撮影者の意図が入っていることが多く、この文献は、これらの操作を自動的に検出する方法について述べている。応用例としては、カメラが動いた量を算出し、その量だけ画像をずらしながら貼り付けることでパノラマを作成するものがある。

【0019】次に、判断205において音楽の終了が判定され、音楽分類処理206において、画像特徴抽出処理204で抽出された特徴量に基づいて音楽が分類される。画像の特徴としてカット点を用いた場合には、カット点が単位時間あたりに出現する頻度を算出し、頻度が高い場合には、音楽のテンポが早く、頻度が低い場合には音楽のテンポが遅いと分類される。また、テロップを用いた場合には、カット点と同様にテロップの出現頻度によって分類するが、表示される文字数が少ない場合には、出現頻度が高くてもテンポが早いとは限らないので、テロップ領域の大きさの逆数を出現頻度に乗じた値を用いる。さらに、被写体あるいはカメラの動きを用いた場合には、動きベクトルが大きいほどテンポが早いと分類される。

【0020】分類情報は、音楽と共に音楽蓄積部103に蓄積され、音楽再生部109によって音楽を検索、再生することが可能である。なお前記において音楽映像の分類について説明したが、当該分類の態様はデータ処理装置が実行できるプログラムの形で保持することができ、本発明は当該プログラムを記録した記録媒体をも含むものである。

【0021】

【発明の効果】(1)請求項1、4および7の発明は、

映像情報を入力し、映像情報の音情報から音楽を検出し、検出された音楽に対応する画像情報の特徴を抽出し、画像の特徴に基づいて音楽を分類することを可能にし、ビートが明確に検出できないような音楽でもテンポによる分類を可能にする。

【0022】(2)請求項2、5および8の発明は、そして請求項3、6および9の発明は、場面の替わり目であるカット点の有無、テロップの有無および大きさ、被写体あるいはカメラの動きのうち、少なくとも1つを検出することにより、音情報を分析する処理を用いずに音楽のテンポを算出することを可能にする。

【図面の簡単な説明】

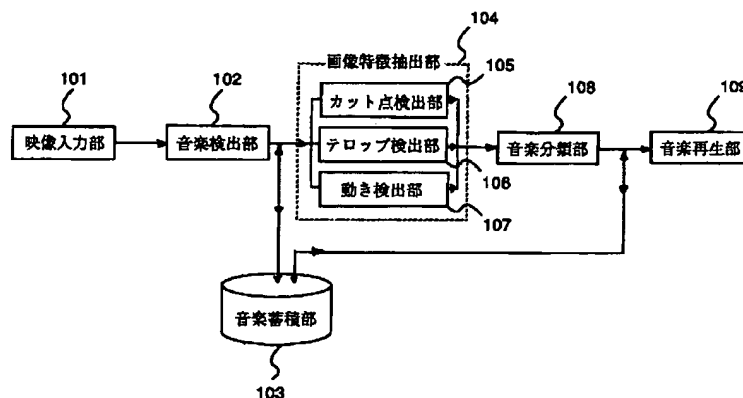
【図1】本発明の一実施形態の音楽映像分類装置の概略構成を示すブロック図である。

【図2】本発明の一実施形態の音楽映像分類装置の処理の流れと本発明をソフトウェアで実現した場合の処理の流れを示すフローチャートである。

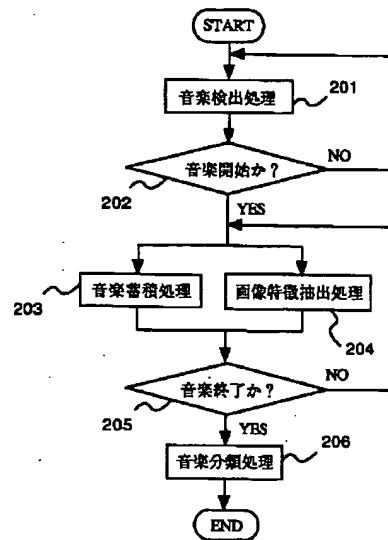
【符号の説明】

- 101 映像入力部
- 102 音楽検出部
- 103 音楽蓄積部
- 104 画像特徴抽出部
- 105 カット点検出部
- 106 テロップ検出部
- 107 動き検出部
- 108 音楽分類部
- 109 音楽再生部
- 201 音楽検出処理
- 202 音楽開始判定処理
- 203 音楽蓄積処理
- 204 画像特徴抽出処理
- 205 音楽終了判定処理
- 206 音楽分類処理

【図1】



【図2】



フロントページの続き

(72)発明者 外村 佳伸  
東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内